Asymptotically Optimal Strategies for Combinatorial Semi-Bandits in Polynomial Time

Thibaut Cuvelier, Richard Combes, Eric Gourdin

Asymptotically Optimal Strategies For Combinatorial Semi-Bandits in Polynomial Time

Thibaut Cuvelier and Richard Combes and Eric Gourdin

February 16, 2021

Abstract

We consider combinatorial semi-bandits with uncorrelated Gaussian rewards. In this article, we propose the first method, to the best of our knowledge, that enables to compute the solution of the Graves-Lai optimization problem in polynomial time for many combinatorial structures of interest. In turn, this immediately yields the first known approach to implement asymptotically optimal algorithms in polynomial time for combinatorial semi-bandits.

1 Introduction

We consider combinatorial bandits, where a learner repeatedly selects decisions x from a combinatorial set $\mathcal{X} \subset \{0,1\}^d$, and obtains random rewards with mean $\theta^\top x$, where θ is an unknown vector. The goal of the learner is to maximize the expected sum of rewards. In the semi-bandit setting, the learner can see several individual components of θ instead of only the total reward $\theta^\top x$.

When decisions in $\mathcal{X}$ have exactly one nonzero entry, the problem reduces to classical bandits [13], for which asymptotically optimal strategies such as KL-UCB and Thompson sampling are known [13, 11, 4]. When $\mathcal{X}$ is a general set, then the problem reduces to linear bandits [8].

Combinatorial semi-bandits have been widely studied. Many authors proposed algorithms and regret upper bounds, including Combinatorial Upper Confidence Bound (CUCB) [12], Efficient Sampling for Combinatorial Bandits (ESCB) [6, 9], Approximate Efficient Sampling for Combinatorial Bandits (AESCB) [7] and a combinatorial version of Thompson sampling (TS) [19]. An information-theoretic regret lower bound was also provided by [6]. Section 2.4 details algorithms and regret guarantees. While our work focuses on stochastic rewards, the adersarial case was also considered, see [1] and references therein.

The main reason why the problem is both interesting and difficult is the combinatorial structure of the decision set $\mathcal{X}$. Many practical problems can be modeled as a combinatorial bandit problem with a particular structure for $\mathcal{X}$, for instance resource allocation (when $\mathcal{X}$ is the set of matchings) or network routing (when $\mathcal{X}$ is a set of source-destination paths in a graph). Typically, $|\mathcal{X}|$ is exponential in the dimension d, so that an exhaustive search over $\mathcal{X}$ is infeasible in practice, and a major challenge is

to derive computationally efficient algorithms. Several authors considered particular structures for $\mathcal{X}$, notably m-sets and matroids, as in this case one can derive stronger results [20, 17, 14]. We will consider more general structures than these two examples.

Combinatorial semi-bandits are a particular case of structured bandits studied by [10, 5]. It is noted that the Graves-Lai regret lower bound [10] generalizes to all structured bandits the well-known Lai-Robbins regret lower bound [13] which holds for classical bandits. For all such problems, there exists asymptotically optimal algorithms such as Optimal Sampling for Structured Bandits (OSSB) [5] under one condition: one must be able to solve a given optimization problem, which we refer to as the Graves-Lai optimization problem. Solving this problem yields both a regret lower bound that holds for any algorithm, as well as an algorithm to attain it, by solving the Graves-Lai problem repeatedly.

Therefore, we believe that one of the most important question to be solved in combinatorial semi-bandits is how to solve the Graves-Lai problem efficiently, in polynomial time in the dimension d. This issue is paramount in solving high-dimensional problems. This is far from straightforward, as the number of variables and constraints in the Graves-Lai problem is proportional to $|\mathcal{X}|$, which is typically exponential in d (see Section 3).

Our contribution We propose the first method, to the best of our knowledge, that enables to compute the solution of the Graves-Lai optimization problem in polynomial time for many combinatorial structures of interest. In turn, this immediately yields the first known approach to implement asymptotically optimal algorithms (such as OSSB) in polynomial time for combinatorial semi-bandits.

The rest of the article is organized as follows. In Section 2, we define the model, give examples of combinatorial structures of interest, and recall the main algorithms for the problem at hand. In Section 3, we introduce the Graves-Lai optimization problem, and show that solving this problem is both necessary and sufficient to obtain asymptotically optimal algorithms. In Section 4, we propose and analyze Graves-Lai Projected Gradient (GLPG), an algorithm to compute the solution to the Graves-Lai optimization problem in polynomial time. Section 5 concludes the paper. Complete proofs are presented in appendix.

2 Model

2.1 Combinatorial Semi-Bandits with Uncorrelated Gaussian Rewards

We consider combinatorial semi-bandits with uncorrelated Gaussian rewards. A learner is given a combinatorial set $\mathcal{X} \subset \{0,1\}^d$, the set of available decisions at each step. Then, for $t = 1, ..., T$, *(i)* the learner chooses a decision $x(t) \in \mathcal{X}$, *(ii)* the environment draws $Y(t) \sim \mathcal{N}(\theta, \frac{1}{2}I_d)$, *(iii)* the learner observes $x(t) \odot Y(t)$ where $\odot$ represents the Hadamard product[1], *(iv)* the learner receives a scalar reward $Y(t)^\top x(t)$. The goal of the learner is to maximize the expected cumulative reward.

[1]The Hadamard product of two vectors x and y in $\mathbb{R}^d$ is the element-wise product: $x \odot y = (x_1 y_1, ..., x_d y_d)$

The vectors $Y(1), ..., Y(t)$ are assumed to be drawn in an i.i.d. fashion from $\mathcal{N}(\theta, \frac{1}{2}I_d)$, so that $Y_1(t), ..., Y_d(t)$ are uncorrelated Gaussian random variables with respective means $\theta_1, ..., \theta_d$ and variance $\frac{1}{2}$. Vector θ is unknown to the learner, and the chosen decision $x(t)$ only depends on $\mathcal{X}$ and the history of observations up to time t, i.e. $(x(1) \odot Y(1)), ..., (x(t-1) \odot Y(t-1))$. In semi-bandit feedback, we observe $x(t) \odot Y(t)$: when $x_i(t) = 1$, we observe $Y_i(t)$, a noisy realization of θ_i that can be used to estimate θ_i. Conversely, when $x_i(t) = 0$, we do not observe anything. Therefore, in order to maximize the reward, we must be able to get accurate estimates of the initially unknown θ; to do so, we must make sure that $x_i(t) = 1$ often enough to get sufficient statistical information about each θ_i.

The goal is to maximize the cumulative reward, or equivalently minimize the total regret. The total regret is defined as the difference in terms of cumulative reward between the learner and that of an oracle who knows θ in hindsight and always selects $x^\star \in \arg\max_{x \in \mathcal{X}} \{\theta^\top x\}$, a decision maximizing the expected reward.

$$R(T, \theta) = T \left(\max_{x \in \mathcal{X}} \{\theta^\top x\} \right) - \sum_{t=1}^{T} \mathbb{E} \left(\theta^\top x(t) \right)$$

The model is summarized in Figure 1.

We introduce some useful notations. For any decision $x \in \mathcal{X}$, we denote by $\Delta_x = \theta^\top x^\star - \theta^\top x$ the reward gap between x and an optimal decision $x^\star$. We define the minimal $\Delta_{\min} = \min_{x \in \mathcal{X} : \Delta_x > 0} \Delta_x$ and maximal reward gap $\Delta_{\max} = \max_{x \in \mathcal{X}} \Delta_x$. Furthermore, we define $m = \max_{x \in \mathcal{X}} \{\mathbf{1}^\top x\}$ the maximal size of a decision, as measured by the number of non-null entries.

Prior knowledge to the learner: combinatorial set $\mathcal{X} \subset \{0, 1\}^d$.
For $t = 1, ..., T$:
1. The learner chooses decision $x(t) \in \mathcal{X}$ based on history
2. The environment draws $Y(t) \sim \mathcal{N}(\theta, \frac{1}{2}I_d)$
3. The learner observes $Y(t) \odot x(t)$ and obtains reward $Y(t)^\top x(t)$

Goal: minimize expected regret: $R(T, \theta) = T(\max_{x \in \mathcal{X}} \{\theta^\top x\}) - \sum_{t=1}^{T} \mathbb{E}(\theta^\top x(t))$

Figure 1: Interaction Between Learner and Environment.

2.2 Combinatorial Structures

Of course, not much can be achieved if the combinatorial set $\mathcal{X}$ is arbitrary. For instance, even when θ is known, if the optimal decision $\max_{x \in \mathcal{X}} \{\theta^\top x\}$ cannot be computed efficiently (e.g., $\mathcal{NP}$-hard), the corresponding combinatorial semi-bandit problem is highly unlikely to have an efficient algorithm. We now highlight the combinatorial structures considered here, which include a large amount of classical and important structures for applications to real-world problems. We consider the same combinatorial structures as [7]. More on combinatorial structures and optimization can

be found in [15] and references therein. For any structure that is defined using a graph $G = (V, E)$, by a slight abuse of notation, we identify a subset of edges with the corresponding vector $x = \{0, 1\}^E$. In that case, the ambient dimension is the number of edges $d = |E|$. Here are the considered combinatorial structures:

- m-sets. Binary vectors with m non-null entries.
- Spanning trees. Spanning trees of a given graph $G = (V, E)$.
- Matroids. Bases of a matroid over a ground set. This includes spanning trees as a particular case.
- Source-destination paths. Paths in a directed, acyclic graph $G = (V, E)$ between a given source and destination.
- Matchings. Matchings in a bipartite graph $G = (V, E)$.
- Intersection of two matroids. Intersection between the sets of bases of two matroids. This includes matchings as a particular case.

2.3 Optimization Problems

As shown below, most if not all of the algorithms for combinatorial semi-bandits involve solving some optimization problems over $\mathcal{X}$. We consider three optimization problems:

- Linear Maximization Compute $\max_{x \in \mathcal{X}} \{a^\top x\}$ (P_{LM})
- Index Maximization Compute $\max_{x \in \mathcal{X}} \{a^\top x + \sqrt{u^\top x}\}$ (P_{IM})
- Budgeted Linear Maximization Compute $\max_{x \in \mathcal{X}} \{a^\top x\}$ subject to $u^\top x \geq s$ (P_{BLM})

where a, u are vectors with positive integer entries and s is a positive scalar. Table 1 indicates whether an algorithm to solve these problems in polynomial time is known. Approximate P_{BLM} means that we can solve P_{BLM} up to a given approximation ratio. The authors in [7] provide algorithms for the polynomial cases depicted in this table. Solving P_{BLM}, either exactly or approximately, is the cornerstone of our approach to design asymptotically optimal algorithms.

	P_{LM}	P_{BLM}	approximate P_{BLM}	P_{IM}
m-sets	✓	✓	✓	✗
spanning trees	✓	✗	✓	✗
matroids	✓	✗	✓	✗
s-t paths	✓	✓	✓	✗
matchings	✓	✗	✓	✗

Table 1: Polynomial Solvability of Combinatorial Problems over $\mathcal{X}$.

2.4 Algorithms, Regret, and Complexity

To understand the interplay between regret and computational efficiency, we now describe the most studied algorithms for combinatorial semi-bandits and highlight their regret guarantees. We define the number of samples obtained up to time t:

$$n_i(t) = \sum_{t'=1}^{t-1} x_i(t), \quad i = 1, ..., d$$

as well as the corresponding empirical mean reward at time t:

$$\hat{\theta}_i(t) = \frac{1}{\max(1, n_i(t))} \sum_{t'=1}^{t-1} x_i(t) Y_i(t), \quad i = 1, ..., d$$

The simplest algorithm is CUCB [12], an extension of the well-known UCB [2] algorithm for stochastic bandits. ESCB [6] is an improved version of CUCB taking advantage of the fact that rewards are not correlated. AESCB [7] is an approximate version of ESCB with lower computational complexity. TS [19] is an algorithm inspired by Bayesian approaches. OSSB [5] is a general, asymptotically optimal algorithm designed for general structured bandits and that can be specialized to combinatorial semi-bandits. These algorithms select a decision $x(t)$ according to the following rules.

- CUCB: $x(t) \in \arg\max_{x \in \mathcal{X}} \{\hat{\theta}(t)^\top x + \sum_{i=1}^d x_i \frac{\ln T}{n_i(t)}\}$
- ESCB: $x(t) \in \arg\max_{x \in \mathcal{X}} \{\hat{\theta}(t)^\top x + \sqrt{\sum_{i=1}^d x_i \frac{\ln T}{n_i(t)}}\}$
- AESCB: $x(t)$ with $\max_{x \in \mathcal{X}} \{\hat{\theta}(t)^\top x + \sqrt{\sum_{i=1}^d x_i \frac{\ln T}{n_i(t)}}\} \leq \hat{\theta}(t)^\top x(t)$ $+\frac{1}{\varepsilon_t} \sqrt{\sum_{i=1}^d x_i(t) \frac{\ln T}{n_i(t)}}$, with δ_t, ε_t two input parameters.
- TS: $x(t) \in \arg\max_{x \in \mathcal{X}} \{V(t)^\top x\}$ where $V(t) \sim \mathcal{N}(\hat{\theta}(t), \mathbf{diag}(\frac{1}{n_1(t)}, ..., \frac{1}{n_d(t)}))$ is a sample from the posterior distribution of θ given the information available at time t

Table 2 summarizes the regret and complexity of algorithms. OSSB is provably asymptotically optimal, while ESCB and AESCB enjoy a $\mathcal{O}(\frac{d(\ln m)^2}{\Delta_{min}} \ln T)$ regret guarantee. TS has a larger regret guarantee of $\mathcal{O}(\frac{d\sqrt{m}}{\Delta_{min}} \ln T)$, and CUCB has the largest one: $\mathcal{O}(\frac{dm}{\Delta_{min}} \ln T)$. There is an interesting interplay here between statistical efficiency (regret) and computational complexity. In terms of complexity, for each time step, CUCB and TS involve solving P_{LM}, while ESCB involves solving P_{IM}, which typically cannot be solved in polynomial time in the dimension d, and AESCB involves solving (up to a fixed approximation ratio) P_{BLM} several times, which can be done in polynomial time. Finally, OSSB involves solving P_{GL}, see section 3.

	Regret	Complexity	Asymptotically Optimal
CUCB	$O(\frac{dm}{\Delta_{\min}} \ln T)$	Solve P_{LM} once	✗
TS	$O(\frac{d\sqrt{m}}{\Delta_{\min}} \ln T)$	Solve P_{LM} once	✗
ESCB	$O(\frac{d(\ln m)^2}{\Delta_{\min}} \ln T)$	Solve P_{IM} once	✗
AESCB	$O(\frac{d(\ln m)^2}{\Delta_{min}} \ln T)$	Approximate P_{BLM} several times	✗
OSSB	$O(C(\theta) \ln T)$	Solve P_{GL} once (see section 3)	✓

Table 2: Algorithms, Regret and Complexity

3 Graves-Lai Formulation: Regret Lower Bound and Asymptotically Optimal Algorithms

Combinatorial semi-bandits are an instance of structured bandits studied in [5], which in turn are an instance of the controlled Markov chains studied by the seminal work of [10]. Those results can be applied to our problem, and in fact doing so yields asymptotically optimal algorithms, i.e. whose regret is, asymptotically, the lowest achievable. We explain how these algorithms function, and explain the challenge of implementing these algorithms efficiently.

3.1 The Graves Lai Optimization Problem

We first introduce the following optimization problem, which we call the Graves-Lai optimization problem for combinatorial semi-bandits:

$$\begin{aligned} \underset{\alpha \in \mathbb{R}_+^{|\mathcal{X}|}}{\text{minimize}} \quad & \sum_{x \in \mathcal{X}} \alpha_x \Delta_x \\ \text{subject to} \quad & \sum_{i \in \mathcal{I}} \frac{x_i}{\sum_{y \in \mathcal{X}} y_i \alpha_y} \leq \Delta_x^2 \ , \ \forall x \in \mathcal{X} \end{aligned} \tag{P_{GL}}$$

with

$$\mathcal{I} = \left\{ i \in \{1, ..., d\} : \max_{x \in \mathcal{X} : x_i = 1} (\theta^\top x) < \max_{x \in \mathcal{X}} (\theta^\top x) \right\}$$

the set of items i that do not appear in any optimal decision. This type of problem first appeared in [10] in the more general context of controlled Markov chains, and was later specialized to combinatorial semi-bandits by [6]. Not only does the Graves-Lai optimization problem yield a regret lower bound that holds for any algorithm, but computing its solution also enables to design algorithms achieving this bound and are hence asymptotically optimal [6].

3.2 Regret Lower Bound

Theorem 1 states that the regret of any uniformly good algorithm (i.e. an algorithm whose regret scales as $o(T^a)$ when $T \to \infty$ for any fixed problem instance and any

$a > 0$) is lower bounded by the optimal value of the Graves-Lai optimization problem. The proof follows from [6][Theorem 1] and is presented in appendix.

Theorem 1 ([10, 6])**.** *Consider a uniformly good algorithm, in the sense that its expected regret verifies $R(T,\theta) = o(T^a)$ for any fixed $\theta \in \mathbb{R}^d$ and $a > 0$.*

Then, its regret verifies for any θ:

$$\liminf_{T\to\infty} \frac{R(T,\theta)}{\ln T} \geq C(\theta)$$

where $C(\theta)$ is the optimal value of the Graves-Lai optimization problem P_{GL}.

The analysis of [10] provides the following interpretation of the objective function and the constraints in P_{GL}. Consider a uniformly good algorithm selecting each sub-optimal decision $x \in \mathcal{X}$ an amount of time equal to $\alpha_x \ln T$. The regret of this algorithm is $(\ln T) \sum_{x\in\mathcal{X}} \alpha_x \Delta_x$, which is proportional to the objective function of P_{GL}. The number of observations to estimate θ_i equals $(\ln T) \sum_{y\in\mathcal{X}} y_i \alpha_y$. Given a sub-optimal decision x, in order to be sure that $x \neq x^\star$, one needs enough statistical information to estimate θ_i for all i such that $x_i = 1$ and $x_i^\star = 0$. To the contrary, if $x_i^\star = 1$, then θ_i can be estimated very accurately without regret, as sampling decision $x^\star$ does not incur regret. More precisely, one can show that the number of observations of any $x \in \mathcal{X}$ must satisfy

$$\sum_{i\in\mathcal{I}} \frac{x_i}{\sum_{y\in\mathcal{X}} y_i \alpha_y} \leq \Delta_x^2$$

Otherwise, it is impossible to distinguish between decision x and the optimal decision $x^\star$ with high probability. In short, the Graves-Lai optimization problem P_{GL} simply consists in minimizing regret, subject to the constraint that one can statistically distinguish between optimal and sub-optimal decisions.

3.3 Asymptotically Optimal Algorithms

In fact, if one can compute the solution of P_{GL}, there exists asymptotically optimal algorithms attaining the lower bound of Theorem 1 such as the doubling trick algorithm of [10] and the arguably simpler OSSB algorithm from [5]. Both of these algorithms are based on certainty equivalence, which involves estimating θ using empirical averages, and selecting each sub-optimal decision an amount of time $\alpha_x^\star \ln T$, where $\alpha^\star$ is an optimal solution of P_{GL}, and where θ is replaced by its estimate. Therefore, the solution of the Graves-Lai optimization problem explicitly gives the way that one should explore sub-optimal decisions to minimize regret.

Theorem 2 ([10, 5])**.** *Assume that one can compute solutions to the Graves-Lai optimization problem for any θ. Then, there exists asymptotically optimal algorithms in the sense that their regret verifies for all θ:*

$$\limsup_{T\to\infty} \frac{R(T,\theta)}{\ln T} \leq C(\theta)$$

3.4 Computational Complexity of Asymptotically Optimal Algorithms

We can conclude that the only difficulty in the design of asymptotically optimal algorithms is a computational one. One must be able to compute solutions to the Graves-Lai optimization problem P_{GL} efficiently. At first look, this seems like a difficult task. Namely, P_{GL} involves optimizing a linear function with $|\mathcal{X}|$ variables, subject to $|\mathcal{X}| + d$ convex constraints. Indeed, for any $x \in \mathcal{X}$, the function

$$\alpha \mapsto \sum_{i \in \mathcal{I}} \frac{x_i}{\sum_{y \in \mathcal{X}} y_i \alpha_y}$$

is convex. Therefore, simply checking whether or not some solution α is feasible may require $\mathcal{O}(|\mathcal{X}|)$ computations, and $\mathcal{O}(|\mathcal{X}|)$ is not polynomial in d for any of the combinatorial structures considered in Section 2.2. Furthermore, even assuming that the optimal solution $\alpha^\star$ can be computed, if the size of $\{x \in \mathcal{X} : \alpha^\star_x > 0\}$ is close to that of $\mathcal{X}$, then simply outputting the optimal solution is not possible in polynomial time. Our main result demonstrates that it is indeed possible to solve P_{GL} in polynomial time, as shown in the next section.

4 Main Result

4.1 Assumptions

Before stating our results, we discuss some of our assumptions.

Assumption 1 (Covering)**.** *For each $i \in \{1, ..., d\}$, there exists a decision $x^i \in \mathcal{X}$ such that $x^i_i = 1$.*

Assumption 1 states that, for all i, there must exist a decision $x^i \in \mathcal{X}$ with $x^i_i = 1$, so that θ_i may be estimated by sampling x^i. If this assumption does not hold, we can simply remove i from consideration, since it plays no role in the Graves-Lai optimization problem. Thus, this assumption can be made without loss of generality.

Assumption 2 (Integrality)**.** *We have that $\theta \in \mathbb{N}^d$.*

Assumption 2 states that the vector θ has positive integer entries. While this makes stating our results simpler, we can easily generalize them to the case where θ has continuous values. Proposition 5 in appendix states that if θ is real valued, we can discretize θ as $\theta^\varepsilon = \varepsilon(\lceil \theta_1/\varepsilon, \rceil, ..., \lceil \theta_d/\varepsilon, \rceil)$ then solve an approximate version of P_{GL} where θ is replaced by $\theta^\varepsilon/\varepsilon$, which has integer entries. This enables us to solve P_{GL} up to an error of $O(1/\varepsilon)$ in time $O\Big(\textbf{poly}(d, 1/\varepsilon)\Big)$. Hence, one can solve P_{GL} up to any fixed accuracy in polynomial time using our results.

Assumption 3 (Polynomial-Time Linear Maximization)**.** *The exact solution of P_{LM} can be computed in time $O\Big(\textbf{poly}(d)\Big)$.*

Assumption 4 (Polynomial-Time Budgeted Linear Maximization)**.** *The exact solution of P_{BLM} can be computed in time $O\Big(\textbf{\textit{poly}}(d, \|u\|_\infty)\Big)$.*

Assumption 5 (Polynomial-Time Approximate Budgeted Linear Maximization)**.** *An ε-optimal solution of P_{BLM} can be computed in time $O\Big(\textbf{\textit{poly}}(d, \|u\|_\infty)\Big)$ for some fixed $\varepsilon > 0$, in the sense that we can compute $\tilde{x} \in \mathcal{X}$ verifying:*

$$a^\top \tilde{x} \geq \varepsilon \left(\max_{x \in \mathcal{X} : u^\top x \geq s} \{a^\top x\} \right) \text{ and } u^\top \tilde{x} \geq s$$

Assumptions 3, 4, and 5 respectively state that one can solve P_{LM} exactly, P_{BLM} exactly, and P_{BLM} approximately. The cases in which those assumptions hold are reported in Table 1. In particular, in all considered combinatorial structures, Assumption 5 does hold, as shown in [7].

Assumption 6 (Compact Representation for Convex Hulls)**.** *The convex hull of $\mathcal{X}$ can be written in the following form:*

$$\textbf{\textit{conv}}(\mathcal{X}) = \{w \in \mathbb{R}^d : Aw = b, w \geq 0\}$$

where the size of A and b is polynomial in the dimension d.

Assumption 6 states that the convex hull of $\mathcal{X}$, a polytope, can be represented in a "compact" manner, i.e. using a polynomial number of linear inequalities. This assumption is verified for all considered combinatorial structures listed above: spanning trees, matchings, paths, etc. (see for instance [15]).

4.2 Main Result

Our main result is Theorem 3. It states that the solution to the Graves-Lai optimization problem can be computed in polynomial time up to any given accuracy. To do so, we design the GLPG (Graves-Lai Projected Gradient) algorithm, which is presented and analyzed below. More precisely, the complexity of GLPG is polynomial in the dimension d, the accuracy level δ and the largest entry in θ, denoted by $\|\theta\|_\infty$. The pseudo-code for GLPG is presented in Figure 2.

Our main result comes in two versions: *(i)* an exact version where one can compute the exact solution up to any given accuracy, when exact Polynomial Time Budgeted Linear Maximization is possible, and *(ii)* an approximate version where one can compute a solution with a fixed approximation ratio up to any given accuracy, when approximate Polynomial Time Budgeted Linear Maximization is possible. If one can only solve the Graves-Lai optimization problem with a fixed approximation ratio, the yielded algorithm is not asymptotically optimal. However, the asymptotic regret of such an algorithm is upper bounded by a universal constant times the Graves-Lai lower bound, which is typically better than what existing algorithms can achieve for large time horizons.

Theorem 3. *Consider $\delta > 0$. Let Assumptions 1, 2, 3 and 6 hold.*

*(Exact version) If Assumption 4 further holds, then the GLPG algorithm outputs α, an δ-optimal solution to P_{GL} in time **poly**$(d, \delta, \|\theta\|_\infty)$ in the sense that:*

$$\sum_{x\in\mathcal{X}} \alpha_x \Delta_x \leq C(\theta) + \delta \text{ and } \sum_{i\in\mathcal{I}} \frac{x_i}{\sum_{y\in\mathcal{X}} y_i \alpha_y} \leq \Delta_x^2 \ \ \forall x \in \mathcal{X},\ \alpha_x \geq 0 \ \ \forall x \in \mathcal{X}$$

*(Approximate version) If Assumption 5 further holds, then the GLPG algorithm outputs α, an (ε, δ)-optimal solution to P_{GL} in time **poly**$(d, \delta, \|\theta\|_\infty)$ in the sense that:*

$$\sum_{x\in\mathcal{X}} \alpha_x \Delta_x \leq \frac{1}{\varepsilon} C(\theta) + \delta \text{ and } \sum_{i\in\mathcal{I}} \frac{x_i}{\sum_{y\in\mathcal{X}} y_i \alpha_y} \leq \Delta_x^2 \ \ \forall x \in \mathcal{X},\ \alpha_x \geq 0 \ \ \forall x \in \mathcal{X}$$

The main steps of the proof are highlighted in the next subsections. We solely prove the approximate version, as the exact version is a particular case of the approximate one with $\varepsilon = 1$.

4.3 Step 0: Computing the set of optimal items

It is noted that $\mathcal{I}$ can be computed in polynomial time using a penalty method. Indeed, one can readily check that $i \in \mathcal{I}$ if and only if

$$\max_{x\in\mathcal{X}} \{\theta^\top y^i\} < \max_{x\in\mathcal{X}} \{\theta^\top x\} \text{ where } y^i \in \arg\max_{x\in\mathcal{X}} \{(\theta + e^i 2d\|\theta\|_\infty)^\top x\} \text{ and } e^i_j = \mathbf{1}\{i = j\}$$

From assumption 3, this computation can be done in polynomial time.

4.4 Step 1: Dimensionality Reduction

The first step in the proof is Proposition 1 proven in appendix. This proposition shows that the solution of P_{GL}, a problem with $|\mathcal{X}|$ variables, can be derived by computing the solution of P'_{GL}, another, much simpler optimization problem with only d variables. The idea behind this reduction is that, instead of optimizing over $(\alpha_x)_{x\in\mathcal{X}}$ (the amount of time each decision is selected), we can optimize over $(\sum_{x\in\mathcal{X}} x_i \alpha_x)_{i=1,\dots,d}$ (the amount of samples obtained to estimate $\theta_1, ..., \theta_d$).

Proposition 1. *Consider $w^\star \in \mathbb{R}^d$ the optimal solution to*

$$\begin{aligned} &\underset{w\in\mathbb{R}^d}{\text{minimize}}\ q^\top w && (P'_{GL}) \\ &\text{subject to } \sum_{i\in\mathcal{I}} \frac{x_i}{w_i} \leq \Delta_x^2 \ \ \forall x \in \mathcal{X}, \quad Mw = 0, w \geq 0, \min_{i\in\mathcal{I}} w_i \geq \underline{w} \end{aligned}$$

where $M \equiv A - \frac{bb^\top A}{\|b\|^2}$ and $q \equiv (\theta^\top x^\star)\frac{b^\top A}{\|b\|^2} - \theta$ and $\underline{w} \equiv (m\|\theta\|_\infty)^{-2}$.

Then there exists $\alpha^\star \in \mathbb{R}^{|\mathcal{X}|}$ an optimal solution to P_{GL} such that: $w^\star = \sum_{x\in\mathcal{X}} x\alpha^\star_x$.

4.5 Step 2: Approximate Subgradient Descent

The next step is to solve the reduced form P'_{GL} using an iterative scheme. To do so, we use a combination of penalization as well as projected subgradient descent. For $x \in \mathcal{X}$, define

$$h_x(w) = \left(\sum_{i \in \mathcal{I}} \frac{x_i}{w_i} \right) - \Delta_x^2$$

as the constraint attached to x in P'_{GL}. Instead of solving P'_{GL}, we solve P''_{GL} in which the constraints are replaced by a penalty, with $\lambda > 0$:

$$\begin{aligned} &\underset{w \in \mathbb{R}^d}{\text{minimize}} \left\{ q^\top w + \lambda \max_{x \in \mathcal{X}} \Big(h_x(w) \Big)^+ \right\} \qquad (P''_{GL}) \\ &\text{subject to } Mw = 0, w \geq 0, \min_{i \in \mathcal{I}} w_i \geq \underline{w}. \end{aligned}$$

where $(\cdot)^+ = \max(\cdot, 0)$ denotes the positive part. The value of λ must be appropriately large to ensure that the constraints in P'_{GL} are satisfied; it will be specified later. Define the polytope

$$\mathcal{M} = \{w : Mw = 0, w \geq 0, \min_{i \in \mathcal{I}} w_i \geq \underline{w}\}$$

We solve P''_{GL} using a strategy that resembles the projected subgradient method. The method is iterative with T iterations [2] and follows the update rule for $t = 1, ..., T$:

$$\begin{aligned} w^0 &= (\underline{w}, ..., \underline{w}) \\ w^{t+1} &= \Pi_{\mathcal{M}} \Big\{ w^t - \eta g^t \Big\}. \\ g^t &= q + \lambda \varepsilon \nabla h_{x^t}(\varepsilon w^t) \mathbf{1}(h_{x^t}(\varepsilon w^t) > 0) \\ \bar{w} &= \frac{1}{T} \sum_{t=1}^{T} w_t. \end{aligned}$$

where x^t is chosen such that

$$\max_{x \in \mathcal{X}} h_x(w^t) \leq h_{x_t}(\varepsilon w^t)$$

and $\Pi_{\mathcal{M}}$ denotes the orthogonal projection on $\mathcal{M}$. The output of the algorithm is the average iterate $\bar{w}$ instead of the last iterate w^T. For some combinatorial sets $\mathcal{X}$, the projection step can be computed exactly in polynomial time; otherwise, it can be computed using an interior point method, a very efficient method for convex optimization programs (see Section 7.8 for more details). In particular, when $\varepsilon = 1$, we have $x^t \in \arg\max_{x \in \mathcal{X}} h_x(w^t)$, so that g_t is simply a subgradient of $w \mapsto \{q^\top w + \lambda \max_{x \in \mathcal{X}} \Big(h_x(w) \Big)^+ \}$ evaluated at w^t and the proposed algorithm follows projected sub-gradient descent for this function. When $\varepsilon < 1$, our algorithm guarantees that, for any x, $h_x(\varepsilon w^t)$ cannot become too large.

[2] In this section t and T denote the iteration number and the total number of iterations of our method. They should not be confused with t and T as defined in the previous sections.

Furthermore, Proposition 2 shows that x^t can be computed in polynomial time under our assumptions, by solving P_{BLM} (exactly or approximately) a polynomial number of times. The proof is in appendix.

Proposition 2. *Under Assumption 2, and either Assumption 4 or 5, x^t can be computed in time **poly**$(d, \delta, \|\theta\|_\infty)$.*

Proposition 3 states that, when λ, η, and T are chosen appropriately, this procedure outputs a solution arbitrarily close to the optimal solution of P'_{GL}. Further, this procedure runs in polynomial time. The proof is involved and is given in appendix.

Proposition 3. *Consider any fixed $\delta > 0$. Let*

$$\begin{aligned}
\delta_2 &= \frac{\delta\varepsilon}{m^2 d\|\theta\|_\infty} \\
\delta_1 &= \frac{\delta}{2(1+\delta_2)} \\
\lambda &= \frac{1}{\delta_2}(\delta_1 + m^2 d\|\theta\|_\infty) \\
T &= \frac{1}{\delta_1^2}\varepsilon^{-2}m^5 d^2\|\theta\|_\infty^2\left(\|q\|^2 + \lambda^2\varepsilon^{-2}dm^8\|\theta\|_\infty^8\right) \\
\eta^2 &= \frac{\varepsilon^{-2}m^5 d^2\|\theta\|_\infty^2}{T(\|q\|^2 + \lambda^2\varepsilon^{-2}dm^8\|\theta\|_\infty^8)}
\end{aligned}$$

Let $\bar{w}$ denote the output of the above procedure, and let $\bar{w}' = (1+\delta_2)\bar{w}$.

Then $\bar{w}'$ is an (ε, δ)-optimal solution to optimization problem P'_{GL} in the sense that

$$\begin{aligned}
&q^\top \bar{w}' \le q^\top(w^\star/\varepsilon) + \delta \\
&M\bar{w}' = 0, \bar{w}' \ge 0, \min_{i\in\mathcal{I}} \bar{w}'_i \ge \underline{w}\,,\ \sum_{i\in\mathcal{I}} \frac{x_i}{\bar{w}'_i} \le \Delta_x^2\,,\ x \in \mathcal{X},
\end{aligned}$$

*and this procedure runs in time **poly**$(d, \delta, \|\theta\|_\infty)$*

4.6 Step 3: Retrieving the Solution to the Original Problem

Assume that we have computed $w^\star \in \mathbb{R}^d$, the optimal solution to P'_{GL}. We now need to retrieve $\alpha^* \in \mathbb{R}^{|\mathcal{X}|}$, the optimal solution to the original problem P_{GL}. Since α^* has $|\mathcal{X}|$ entries, and $|\mathcal{X}|$ is typically not polynomial in the dimension d, this seems like an impossible task. However, we can choose α^* such that most of its entries are zero: the optimum solution is generally not unique. From Carathéodory's theorem, any point in the convex hull of $|\mathcal{X}|$ can be written as a convex combination of at most $d+1$ elements of $\mathcal{X}$.

We provide an iterative procedure to compute α^* knowing $w^\star$ and analyze it in Proposition 4. We let $\bar{w}^1 = w^\star$ and for $k = 1, ..., d$: if $\bar{w}^k = 0$, we let $\alpha_{x^k} = 0$ and $x^k \in \mathcal{X}$ chosen arbitrarily; otherwise, we let $\alpha_{x^k} = \min_{i:\bar{w}_i^k > 0} \bar{w}_i^k$ and $\bar{w}^{k+1} = \bar{w}^k - \alpha_{x^k} x^k$ where

$$x^k \in \arg\min_{x\in\mathcal{X}} \Big\{ \sum_{i=1}^d x_i \mathbf{1}\{\bar{w}_i^k > 0\} \Big\}$$

The output of this procedure is $\alpha_{x^1}, ..., \alpha_{x^d}$ and $x^1, ..., x^d$, which is a decomposition of $w^\star$ as a linear combination with positive coefficients with at most d elements from $\mathcal{X}$. The cornerstone of this procedure is the fact that if $\bar{w} = \sum_{x \in \mathcal{X}} x\alpha_x$ with $\alpha \geq 0$ then for any x such that $\alpha_x > 0$ we have that $x_i = 1$ implies $\bar{w}_i > 0$. This concludes the proof of Theorem 3.

Proposition 4. *The above procedure is such that* $w^\star = \sum_{k=1}^{d} x^k \alpha_{x^k}$ *with* $\alpha_{x^1}, ..., \alpha_{x^d}$ *positive numbers and runs in time* ***poly***$(d, \delta, \|\theta\|_\infty)$.

5 Conclusion

We have proposed the first method, to the best of our knowledge, to compute the solution of the Graves-Lai optimization problem for combinatorial semi-bandits in polynomial time, which in turn allows to implement asymptotically optimal algorithms (such as OSSB) for this problem. Our results hold for a large number of combinatorial structures including m-sets, spanning trees, paths, and matchings. We believe that our results shed some light on the trade-off between statistical efficiency and computational complexity in bandit optimization.

References

[1] Jean-Yves Audibert, Sébastien Bubeck, and Gábor Lugosi. Regret in online combinatorial optimization. Mathematics of Operations Research, 39(1):31–45, 2013.

[2] Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. Finite-time Analysis of the Multiarmed Bandit Problem. Mach. Learn., 47(2-3):235–256, May 2002.

[3] Stephen Boyd and Lieven Vandenberghe. Convex Optimization. Cambridge University Press, USA, 2004.

[4] O. Cappé, A. Garivier, O. Maillard, R. Munos, and G. Stoltz. Kullback-leibler upper confidence bounds for optimal sequential allocation. Annals of Statistics, 41(3):516–541, June 2013.

[5] Richard Combes, Stefan Magureanu, and Alexandre Proutiere. Minimal exploration in structured stochastic bandits. In Proc. of NIPS, 2017.

[6] Richard Combes, Sadegh Talebi, Alexandre Proutière, and Marc Lelarge. Combinatorial Bandits Revisited. In Proc. of NIPS, 2015.

[7] Thibaut Cuvelier, Richard Combes, and Eric Gourdin. Statistically efficient, polynomial time algorithms for combinatorial semi bandits. In Proc. of ACM SIGMETRICS, 2021.

[8] V. Dani, T. P. Hayes, and S. M. Kakade. Stochastic linear optimization under bandit feedback. In Proc. of COLT, 2008.

[9] Remy Degenne and Vianney Perchet. Combinatorial semi-bandit with known covariance. In Proc. of NIPS, 2016.

[10] Todd L. Graves and Tze Leung Lai. Asymptotically efficient adaptive choice of control laws in controlled markov chains. SIAM Journal on Control and Optimization, 35(3):715–743, 1997.

[11] Emilie Kaufmann, Nathaniel Korda, and Rémi Munos. Thompson sampling: An asymptotically optimal finite-time analysis. In Proc. of ALT, 2012.

[12] Branislav Kveton, Zheng Wen, Azin Ashkan, and Csaba Szepesvari. Tight regret bounds for stochastic combinatorial semi-bandits. In Proc. of AISTATS, 2015.

[13] T.L. Lai and H. Robbins. Asymptotically efficient adaptive allocation rules. Advances in Applied Mathematics, 6(1):4–2, 1985.

[14] Pierre Perrault, Vianney Perchet, and Michal Valko. Exploiting structure of uncertainty for efficient matroid semi-bandits. In Proc. of ICML, 2019.

[15] A. Schrijver. Combinatorial Optimization - Polyhedra and Efficiency. Springer, 2003.

[16] Shai Shalev-Shwartz and Shai Ben-David. Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press, USA, 2014.

[17] Mohammad Sadegh Talebi and Alexandre Proutiere. An optimal algorithm for stochastic matroid bandit optimization. In Proc. of ICAAMS, 2016.

[18] Fei Wang, Ping Li, Arnd Christian König, and Arnd Christian König. Learning a bi-stochastic data similarity matrix. In Proc. of IEEE ICDM, 2010.

[19] Siwei Wang and Wei Chen. Thompson sampling for combinatorial semi-bandits. In Proc. of ICML, 2018.

[20] Zheng Wen, Branislav Kveton, and Azin Ashkan. Efficient learning in large-scale combinatorial semi-bandits. In Proc. of ICML, 2015.

6 Additional Result

To avoid confusion, in this section, for any parameter $\lambda \in \mathbb{R}^d$, we use the notation

$$\Delta_x(\lambda) = \max_{y \in \mathcal{X}} (\lambda^\top y) - \lambda^\top x$$

to denote the reward gap of decision $x \in \mathcal{X}$ under parameter λ.

Proposition 5. *Consider $\varepsilon > 0$ and a real valued vector $\theta = \mathbb{R}^d$. Define the discretized vector*

$$\theta^\varepsilon = \varepsilon(\lceil \theta_1/\varepsilon, \rceil, ..., \lceil \theta_d/\varepsilon, \rceil)$$

Consider the following optimization problem which approximates P_{GL}:

$$\begin{aligned} &\underset{\alpha \in \mathbb{R}_+^{|\mathcal{X}|}}{\textit{minimize}} \sum_{x \in \mathcal{X}^\star} \alpha_x \Delta_x(\theta^\varepsilon) && (P^\varepsilon_{GL}) \\ &\textit{subject to} \sum_{i \in \mathcal{I}} \frac{x_i}{\sum_{y \in \mathcal{X}} y_i \alpha_y} \leq (\Delta_x(\theta^\varepsilon))^2 \ , \ \forall x \in \mathcal{X}^\star \end{aligned}$$

with

$$\begin{aligned} \mathcal{I} &= \left\{ i \in \{1, ..., d\} : \max_{x \in \mathcal{X}: x_i = 1} (\theta^\top x) < \max_{x \in \mathcal{X}} (\theta^\top x) \right\} \\ \mathcal{X}^\star &= \{x \in \mathcal{X} : \Delta_x(\theta) > 0\} \end{aligned}$$

Denote by $\alpha^{\star,\varepsilon}$ an optimal solution to P^ε_{GL} and $\alpha^\star$ an optimal solution to P_{GL}. Assume that $\varepsilon \leq \frac{\Delta_{\min}}{2}$. Then $\alpha^{\star,\varepsilon}(1 + \frac{2m\varepsilon}{\Delta_{\min}})^2$ is a feasible solution to P_{GL}, and it is near optimal in the sense that:

$$\sum_{x \in \mathcal{X}^\star} \alpha^{\star,\varepsilon}_x \Delta_x(\theta) \leq (1 + \frac{4m\varepsilon}{\Delta_{\min}})^4 \sum_{x \in \mathcal{X}^\star} \alpha^\star_x \Delta_x(\theta)$$

Proof: We first upper bound the gap differences. For any x we have

$$\begin{aligned} |\Delta_x(\theta) - \Delta_x(\theta^\varepsilon)| &\leq |\max_{y \in \mathcal{X}} (\theta^\top y) - \max_{y \in \mathcal{X}} ((\theta^\varepsilon)^\top y)| + |\theta^\top x - (\theta^\varepsilon)^\top x| \\ &\leq \max_{y \in \mathcal{X}} |(\theta - \theta^\varepsilon)^\top y| + |(\theta - \theta^\varepsilon)^\top x| \\ &\leq 2 \max_{y \in \mathcal{X}} |(\theta - \theta^\varepsilon)^\top y| \\ &\leq 2 \|\theta - \theta^\varepsilon\|_\infty \max_{y \in \mathcal{X}} (\mathbf{1}^\top y) \\ &\leq 2m\varepsilon \end{aligned}$$

Based on the above inequality, for any $x \in \mathcal{X}^\star$ we have:

$$\begin{aligned} \Delta_x(\theta^\varepsilon) = \Delta_x(\theta) \frac{\Delta_x(\theta^\varepsilon)}{\Delta_x(\theta)} &\leq \Delta_x(\theta) \frac{\Delta_x(\theta) + 2m\varepsilon}{\Delta_x(\theta)} \\ &= \Delta_x(\theta)(1 + \frac{2m\varepsilon}{\Delta_x(\theta)}) \leq \Delta_x(\theta)(1 + \frac{2m\varepsilon}{\Delta_{\min}}) \end{aligned}$$

Similarly, for any $x \in \mathcal{X}^\star$ we have:

$$\begin{aligned}\Delta_x(\theta) &= \Delta_x(\theta^\varepsilon)\frac{\Delta_x(\theta)}{\Delta_x(\theta^\varepsilon)} \leq \Delta_x(\theta^\varepsilon)\frac{\Delta_x(\theta^\varepsilon)+2m\varepsilon}{\Delta_x(\theta^\varepsilon)} \\ &\leq \Delta_x(\theta^\varepsilon)(1+\frac{2m\varepsilon}{\Delta_x(\theta^\varepsilon)}) \leq \Delta_x(\theta^\varepsilon)(1+\frac{4m\varepsilon}{\Delta_{\min}})\end{aligned}$$

where we used the fact that

$$\Delta_x(\theta^\varepsilon) \geq \Delta_x(\theta) - 2m\varepsilon \geq \frac{\Delta_{\min}}{2}$$

since $\varepsilon \leq \frac{\Delta_{\min}}{2m}$.

We now turn to the relationship between P_{GL} and P^ε_{GL}. Consider $x \in \mathcal{X}^\star$. Since $\alpha^{\star,\varepsilon}$ is an optimal solution to P^ε_{GL}, we must have

$$\sum_{i\in\mathcal{I}} \frac{x_i}{\sum_{y\in\mathcal{X}} y_i \alpha_y^{\star,\varepsilon}} \leq (\Delta_x(\theta^\varepsilon))^2 \leq (\Delta_x(\theta))^2(1+\frac{2m\varepsilon}{\Delta_{\min}})^2$$

using our previous reasoning. We have proven that $\alpha^{\star,\varepsilon}(1+\frac{2m\varepsilon}{\Delta_{\min}})^2$ is a feasible solution to P_{GL}.

Using the same technique, since $\alpha^\star$ is an optimal solution to P_{GL}, we have that

$$\sum_{i\in\mathcal{I}} \frac{x_i}{\sum_{y\in\mathcal{X}} y_i \alpha_y^{\star}} \leq (\Delta_x(\theta))^2 \leq (\Delta_x(\theta^\varepsilon))^2(1+\frac{4m\varepsilon}{\Delta_{\min}})^2$$

We have proven that $\alpha^\star(1+\frac{4m\varepsilon}{\Delta_{\min}})^2$ is a feasible solution to P^ε_{GL}.

We can now conclude:

$$\begin{aligned}\sum_{x\in\mathcal{X}^\star} \alpha^{\star,\varepsilon}\Delta_x(\theta) &\leq (1+\frac{4m\varepsilon}{\Delta_{\min}}) \sum_{x\in\mathcal{X}^\star} \alpha^{\star,\varepsilon}\Delta_x(\theta^\varepsilon) \\ &\leq (1+\frac{4m\varepsilon}{\Delta_{\min}})^3 \sum_{x\in\mathcal{X}^\star} \alpha^{\star}\Delta_x(\theta^\varepsilon) \\ &\leq (1+\frac{4m\varepsilon}{\Delta_{\min}})^4 \sum_{x\in\mathcal{X}^\star} \alpha^{\star}\Delta_x(\theta)\end{aligned}$$

where we successively used the inequality derived above, the fact that $\alpha^\star(1+\frac{4m\varepsilon}{\Delta_{\min}})^2$ is a feasible solution to P^ε_{GL} and the inequality derived above again.

7 Proofs

7.1 Proof of Theorem 1

From [6][Theorem 1], the result holds when $C(\theta)$ is the value of the following optimization problem:

$$\begin{aligned} &\underset{\alpha \in \mathbb{R}_+^{|\mathcal{X}|}}{\text{minimize}} \sum_{x \in \mathcal{X}} \alpha_x \Delta_x \\ &\text{subject to } \min_{\lambda \in B(\theta)} \left\{ \sum_{i=1}^{d} \sum_{x \in \mathcal{X}} \alpha_x x_i D(\theta_i, \lambda_i) \right\} \geq 1 \end{aligned}$$

where

$$B(\theta) = \left\{ \lambda \in \mathbb{R}^d : \lambda^\top x^\star < \max_{x \in \mathcal{X}} \{\lambda^\top x\} \text{ and } \theta_i = \lambda_i \ , \ \forall i \notin \mathcal{I} \right\}$$

is the set of parameters λ under which $x^\star$ is not the optimal decision, and such that λ cannot be distinguished from θ when selecting only optimal decisions under θ. $D(\theta_i, \lambda_i)$ is the Kullback Leibler divergence between the distribution of the rewards for i with respective means θ_i and λ_i. Since rewards are Gaussian with variance $\frac{1}{2}$, the divergence is given by $D(\theta_i, \lambda_i) = (\theta_i - \lambda_i)^2$. Furthermore, if $i \notin \mathcal{I}$, then $\theta_i = \lambda_i$, so that $D(\theta_i, \lambda_i) = 0$. Thus, the optimization problem simplifies to:

$$\begin{aligned} &\underset{\alpha \in \mathbb{R}_+^{|\mathcal{X}|}}{\text{minimize}} \sum_{x \in \mathcal{X}} \alpha_x \Delta_x \\ &\text{subject to } \min_{\lambda \in B(\theta)} \left\{ \sum_{i \in \mathcal{I}} \sum_{x \in \mathcal{X}} \alpha_x x_i (\theta_i - \lambda_i)^2 \right\} \geq 1 \end{aligned}$$

Decompose $B(\theta)$ according to the optimal decision and its value as follows:

$$\begin{aligned} B(\theta) &= \cup_{v>0} \cup_{x \neq x^\star} B_{x,v}(\theta) \\ B_{x,v}(\theta) &= \{\lambda \in B(\theta) : \lambda^\top x = \theta^\top x^\star + v\} \end{aligned}$$

Thus, the optimum solution of

$$\min_{\lambda \in B_{x,v}(\theta)} \left\{ \sum_{i \in \mathcal{I}} \sum_{x \in \mathcal{X}} \alpha_x x_i (\theta_i - \lambda_i)^2 \right\}$$

is $\lambda \in \mathbb{R}^d$ minimizing the quadratic function $\sum_{i \in \mathcal{I}} \sum_{x \in \mathcal{X}} \alpha_x x_i (\theta_i - \lambda_i)^2$ subject to the linear equality constraint $\lambda^\top x = \theta^\top x^\star + v$. Writing the Karush-Kuhn-Tucker conditions and solving, we can check that the minimum is:

$$\min_{\lambda \in B_{x,v}(\theta)} \left\{ \sum_{i \in \mathcal{I}} \sum_{x \in \mathcal{X}} \alpha_x x_i (\theta_i - \lambda_i)^2 \right\} = \frac{(\Delta_x + v)^2}{\sum_{i \in \mathcal{I}} x_i (\sum_{y \in \mathcal{X}} y_i \alpha_y)^{-1}}$$

The constraint

$$\min_{\lambda \in B(\theta)} \left\{ \sum_{i \in \mathcal{I}} \sum_{x \in \mathcal{X}} \alpha_x x_i (\theta_i - \lambda_i)^2 \right\} \geq 1$$

is satisfied if and only if the above is greater than 1 for all $x \in \mathcal{X}$ and all $v > 0$, i.e:

$$\sum_{i \in \mathcal{I}} \frac{x_i}{\sum_{y \in \mathcal{X}} y_i \alpha_y} \leq \Delta_x^2 \ , \ \forall x \in \mathcal{X}$$

Therefore, the original optimization problem is, as claimed, the Graves-Lai optimization problem

$$\begin{aligned} &\underset{\alpha \in \mathbb{R}_+^{|\mathcal{X}|}}{\text{minimize}} \ \sum_{x \in \mathcal{X}} \alpha_x \Delta_x && (P_{GL}) \\ &\text{subject to} \ \sum_{i \in \mathcal{I}} \frac{x_i}{\sum_{y \in \mathcal{X}} y_i \alpha_y} \leq \Delta_x^2 \ , \ \forall x \in \mathcal{X} \end{aligned}$$

This concludes the proof.

7.2 Proof of Proposition 1

We start by stating the definition of P_{GL}, and notice that both the objective function and the constraints solely depend on $\sum_x \alpha_x$ and $\sum_x x\alpha_x$.

$$\begin{aligned} &\underset{\alpha \in \mathbb{R}_+^{|\mathcal{X}|}}{\text{minimize}} \ \sum_{x \in \mathcal{X}} \alpha_x \Delta_x \qquad (P_{GL}) \\ &\text{subject to} \ \sum_{i \in \mathcal{I}} \frac{x_i}{\sum_{y \in \mathcal{X}} y_i \alpha_y} \leq \Delta_x^2, x \in \mathcal{X} \end{aligned}$$

Those variables live in the following set:

$$\begin{aligned} \left\{ \left(\sum_{x \in \mathcal{X}} x\alpha_x, \sum_{x \in \mathcal{X}} \alpha_x \right) : \alpha \in \mathbb{R}_+^{|\mathcal{X}|} \right\} &= \{(w, v) : \frac{w}{v} \in \textbf{conv}(\mathcal{X}), v \geq 0\} \\ &= \left\{ (w, v) : A\frac{w}{v} = b, \frac{w}{v} \geq 0, v \geq 0 \right\} \\ &= \{(w, v) : Aw = vb, w \geq 0, v \geq 0\} . \end{aligned}$$

where $\textbf{conv}(\mathcal{X})$ denoted the convex hull of $\mathcal{X}$, and we have used Assumption 6.

If $b = 0$, we simply have that $Aw = 0$; otherwise, $Aw = vb$. Therefore, $b^\top Aw = vb^\top b = v\|b\|^2$ and $v = \frac{b^\top Aw}{\|b\|^2}$. This implies that $Aw = vb$ if and only if

$$0 = Aw - vb = Aw - \frac{bb^\top Aw}{\|b\|^2} = Mw$$

by definition of M. Therefore,

$$\begin{aligned}\sum_{x\in\mathcal{X}} \alpha_x \Delta_x &= \left(\theta^\top x^\star\right)\left(\sum_{x\in\mathcal{X}} \alpha_x\right) - \theta^\top \left(\sum_{x\in\mathcal{X}} x\alpha_x\right) \\ &= (\theta^\top x^\star) v + \theta^\top w \\ &= \left(\theta^\top x^\star\right) \frac{b^\top A w}{\|b\|^2} - \theta^\top w \\ &= \left(\left(\theta^\top x^\star\right) \frac{b^\top A}{\|b\|^2} - \theta\right)^\top w \\ &= q^\top w\end{aligned}$$

by definition of q.

By Assumption 1, for any $i \in \mathcal{I}$, there exists x^i such that $x^i_i = 1$. As a consequence, for any feasible solution w,

$$\frac{1}{w_i} \leq \sum_{j\in\mathcal{I}} \frac{x^i_j}{w_j} \leq \Delta^2_{x^i} \leq (\theta^\top x^\star)^2 \leq m^2 \|\theta\|^2_\infty$$

Thus, we can impose the additional constraint that $\min_{i\in\mathcal{I}} w_i \geq \underline{w} \equiv (m\|\theta\|_\infty)^{-2}$ for $i \in \mathcal{I}$.

This yields the claimed reduced form:

$$\begin{aligned}&\underset{w\in\mathbb{R}^d}{\text{minimize}}\ q^\top w \qquad (P'_{GL}) \\ &\text{subject to } \sum_{i\in\mathcal{I}} \frac{x_i}{w_i} \leq \Delta^2_x, x \in \mathcal{X}, \quad Mw = 0, w \geq 0, \min_{i\in\mathcal{I}} w_i \geq \underline{w},\end{aligned}$$

which concludes the proof of the proposition.

7.3 Technical Lemma: Optimal Solution

Lemma 1. *Define $\alpha^\star_x$ an optimal solution to P_{GL}. Define $w^\star = \sum_{x\in\mathcal{X}} x\alpha^\star_x$ the corresponding solution to P'_{GL}.*

Then, its value is upper bounded by

$$q^\top w^\star = \sum_{x\in\mathcal{X}} \alpha^\star_x \Delta_x \leq md\frac{\Delta_{\max}}{\Delta^2_{\min}}$$

and the norm of the optimal solution is upper bounded by

$$\|w^\star\| = \left\|\sum_{x\in\mathcal{X}} x\alpha^\star_x\right\| \leq m^{\frac{3}{2}} d\frac{\Delta_{\max}}{\Delta^2_{\min}}$$

Furthermore, if Assumption 2 holds, we have

$$q^\top w^\star \leq m^2 d\|\theta\|_\infty$$

and

$$\|w^\star\| \leq m^{\frac{5}{2}} d \|\theta\|_\infty$$

Define $w^\star = \sum_{x \in \mathcal{X}} x\alpha_x$. From Assumption 1, for each $i \in \mathcal{I}$, consider $x^i \in \mathcal{X}$ such that $x^i_i = 1$. Consider

$$w = \sum_{i=1}^{d} \frac{m}{\Delta^2_{\min}} x^i$$

This implies that for all $i = 1, ..., d$

$$w_i \geq \frac{m}{\Delta^2_{\min}}$$

and in turn for any x:

$$\sum_{i \in \mathcal{I}} \frac{x_i}{w_i} \leq \Delta^2_{\min} \frac{1}{m} \sum_{i \in \mathcal{I}} x_i \leq \Delta^2_{\min} \frac{1}{m} (\sum_{i=1}^{d} x_i) \leq \Delta^2_{\min} \leq \Delta^2_x$$

Hence, w is a feasible solution, which implies that

$$q^\top w^\star \leq q^\top w = \sum_{i=1}^{d} m \frac{\Delta_{x^i}}{\Delta^2_{\min}} \leq md \frac{\Delta_{\max}}{\Delta^2_{\min}}$$

Now, by definition the optimal solution can be expressed as:

$$w^\star = \sum_{x \in \mathcal{X}} x\alpha_x$$

First, notice that

$$q^\top w^\star = \sum_{x \in \mathcal{X}} \alpha_x \Delta_x \geq \Delta_{\min} \sum_{x \in \mathcal{X}} \alpha_x$$

Consequently,

$$\sum_{x \in \mathcal{X}} \alpha_x \leq \frac{q^\top w^\star}{\Delta_{\min}}$$

Using the triangle inequality:

$$\left\| \sum_{x \in \mathcal{X}} x\alpha_x \right\| \leq \sum_{x \in \mathcal{X}} \alpha_x \|x\| \leq \sqrt{m} \sum_{x \in \mathcal{X}} \alpha_x \leq m \frac{q^\top w^\star}{\Delta_{\min}} \leq m^2 d \frac{\Delta_{\max}}{\Delta^2_{\min}}$$

This proves the first result.

If Assumption 2 holds as well, we have

$$1 \leq \Delta_{\min} \leq \Delta_{\max} \leq m\|\theta\|_\infty$$

which proves the second result.

7.4 Proof of Proposition 2

From Assumption 2, θ has positive integer components: for any $x \in \mathcal{X}$, we have $\theta^\top x \in \{0, ..., m\|\theta\|_\infty\}$. In turn, this implies that $\Delta_x \in \{0, ..., m\|\theta\|_\infty\}$ for all $x \in \mathcal{X}$. Now, we use Assumption 4 or 5 to compute (in polynomial time), for $s \in \{0, ..., m\|\theta\|_\infty\}$, an ε-approximate solution to P_{BLM} denoted by $X^{t,s} \in \mathcal{X}$ with

$$\sum_{i\in\mathcal{I}} \frac{X_i^{t,s}}{w_i^t} \geq \varepsilon \left(\max_{x\in\mathcal{X}:\Delta_x\leq s} \sum_{i\in\mathcal{I}} \frac{X_i^{t,s}}{w_i^t} \right) \text{ and } \Delta_x \leq s$$

and one may readily check that

$$x^t = X^{t,s^t} \text{ with } s^t \in \arg\max_{s\in\{0,...,m\|\theta\|_\infty\}} \left\{ \sum_{i\in\mathcal{I}} \frac{X_i^{t,s}}{w_i^t} - s^2 \right\}$$

satisfies $\max_{x\in\mathcal{X}} h_x(w^t) \leq h_{x^t}(\varepsilon w^t)$. In summary, x^t can be computed in time $O(\textbf{poly}(d, \delta, \|\theta\|_\infty))$.

7.5 Technical Lemma: Gradient Descent

We first state a technical lemma due to [16][Lemma 14.1].

Lemma 2. *Consider $\mathcal{M}$ a convex set, $\eta > 0$, $\hat{w}$ and $g^1, ..., g^T$ arbitrary vectors, w^n a sequence defined as*

$$w^{t+1} = \Pi_{\mathcal{M}}\{w^t - \eta g^t\}$$

with $\Pi_{\mathcal{M}}$ the orthogonal projection onto $\mathcal{M}$.

Then we have:

$$\sum_{t=1}^{T} \langle w^t - \hat{w}, g^t \rangle \leq \frac{\|\hat{w}\|^2}{2\eta} + \frac{\eta}{2} \sum_{t=1}^{T} \|v^t\|^2$$

[16][Lemma 14.1] first states the lemma without the projection step, and afterwards argue that their proof still holds when a projection step is added, which corresponds to Lemma 2.

7.6 Proof of Proposition 3

The procedure 2 runs in time $O(\textbf{poly}(d, \delta, \|\theta\|_\infty))$. Indeed, the number of iterations is $T = O(\textbf{poly}(d, \delta, \|\theta\|_\infty))$, and each iteration takes time $O(\textbf{poly}(d, \delta, \|\theta\|_\infty))$.

Define the error:

$$E = q^\top \bar{w} - q^\top (w^\star/\varepsilon) + \lambda \max_{x\in\mathcal{X}} (h_x(\bar{w}))^+$$

Using Jensen's inequality, since $w \mapsto \max_{x\in\mathcal{X}} (h_x(w))^+$ is convex,

$$E \leq \frac{1}{T} \sum_{t=1}^{T} (q^\top w^t) - q^\top (t^\star/\varepsilon) + \lambda \frac{1}{T} \sum_{t=1}^{T} \max_{x\in\mathcal{X}} (h_x(w^t))^+$$

We use the following notation for the dot product:

$$q^\top w^t - q^\top (w^\star/\varepsilon) = \langle w^t - w^\star/\varepsilon, q\rangle$$

By definition of x^t:

$$\max_{x\in\mathcal{X}} (h_x(w^t))^+ \leq (h_{x^t}(\varepsilon w^t))^+$$

Using the fact that $w \mapsto (h_{x^t}(\varepsilon w))^+$ is a convex function and one of its subgradients is

$$\varepsilon \nabla h_{x^t}(\varepsilon w)\mathbf{1}(h_{x^t}(\varepsilon w) > 0)$$

we get, by definition of a subgradient:

$$(h_{x^t}(\varepsilon w))^+ - (h_{x^t}(\varepsilon(w^\star/\varepsilon)))^+ \leq \langle w^t - w^\star/\varepsilon, \varepsilon\nabla h_{x^t}(\varepsilon w^t)\mathbf{1}(h_{x^t}(\varepsilon w^t) > 0)\rangle$$

We have that $h_x(w^\star) < 0$ for all x by definition of $w^\star$, so that $(h_{x^t}(\varepsilon(w^\star/\varepsilon)))^+ = 0$ and replacing above we get:

$$\max_{x\in\mathcal{X}} (h_x(w^t))^+ \leq \langle w^t - w^\star/\varepsilon, \varepsilon\nabla h_{x^t}(\varepsilon w^t)\mathbf{1}(\nabla h_{x^t}(\varepsilon w^t) > 0)\rangle$$

Thus:

$$E \leq \frac{1}{T}\sum_{t=1}^{T} \langle w^t - w^\star/\varepsilon, q + \lambda\varepsilon\nabla h_{x^t}(\varepsilon w^t)\mathbf{1}(h_{x^t}(\varepsilon w^t) > 0)\rangle$$

Using the fact that:

$$\begin{aligned} w^{t+1} &= \Pi_{\mathcal{M}}\Big\{w^t - \eta g^t\Big\}. \\ g^t &= q + \lambda\varepsilon\nabla h_{x^t}(\varepsilon w^t)\mathbf{1}(h_{x^t}(\varepsilon w^t) > 0) \\ \bar{w} &= \frac{1}{T}\sum_{t=1}^{T} w_t. \end{aligned}$$

Lemma 2 yields:

$$E \leq \frac{1}{2T}\left(\frac{\|w^0 - w^\star/\varepsilon\|^2}{\eta} + \eta\sum_{t=1}^{T}\|q + \lambda\varepsilon\nabla h_{x^t}(\varepsilon w^t)\mathbf{1}(h_{x^t}(\varepsilon w^t) > 0)\|^2\right)$$

so that

$$E \leq \frac{1}{2T}\left(\frac{\|w^0 - w^\star/\varepsilon\|^2}{\eta} + \eta\sum_{t=1}^{T}(\|q\|^2 + (\lambda\varepsilon)^2\|\nabla h_{x^t}(\varepsilon w^t)\|^2)\right)$$

We may upper bound each term in the expression above as follows.

Since

$$\nabla h_{x^t}(\varepsilon w^t) = -\left(\frac{x_1^t}{(\varepsilon w_1)^2}, ..., \frac{x_d^t}{(\varepsilon w_d^t)^2}\right),$$

the gradient term is upper bounded as

$$\|\nabla h_{x^t}(\varepsilon w^t)\|^2 = \sum_{i\in\mathcal{I}} \frac{x_i^t}{(\varepsilon w_i^t)^4} \leq \frac{d}{(\varepsilon \underline{w})^4} = \varepsilon^{-4} d m^8 \|\theta\|_\infty^8.$$

where we used the fact that $w^t \in \mathcal{M}$, which implies $w^t \geq \underline{w}$.

Since $w^0 = (\underline{w}, ..., \underline{w})$ and $w^\star/\varepsilon \geq \underline{w}$, we get:

$$\|w^0 - w^\star/\varepsilon\| \leq \varepsilon^{-1}\|w^\star\| \leq \varepsilon^{-1} m^{\frac{5}{2}} d \|\theta\|_\infty$$

using Lemma 1.

Replacing, we get the upper bound:

$$E \leq \frac{1}{2T}\left(\frac{\varepsilon^{-2} m^5 d^2 \|\theta\|_\infty^2}{\eta} + \eta T\left(\|q\|^2 + \lambda^2 \varepsilon^{-2} d m^8 \|\theta\|_\infty^8\right)\right)$$

setting η to equalize both terms

$$\eta^2 = \frac{\varepsilon^{-2} m^5 d^2 \|\theta\|_\infty^2}{T(\|q\|^2 + \lambda^2 \varepsilon^{-2} d m^8 \|\theta\|_\infty^8)}$$

so that the optimization error has the following upper bound:

$$E \leq \frac{\varepsilon^{-2} m^5 d^2 \|\theta\|_\infty^2}{\eta T} = \frac{1}{\sqrt{T}} \varepsilon^{-1} m^{5/2} d \|\theta\|_\infty \sqrt{\|q\|^2 + \lambda^2 \varepsilon^{-2} d m^8 \|\theta\|_\infty^8}$$

Recall the definitions:

$$\delta_2 = \frac{\delta\varepsilon}{md\|\theta\|_\infty}$$

and

$$\delta_1 = \frac{\delta}{2(1+\delta_2)}$$

Now, setting

$$T = \frac{1}{\delta_1^2} \varepsilon^{-2} m^5 d^2 \|\theta\|_\infty^2 \left(\|q\|^2 + \lambda^2 \varepsilon^{-2} d m^8 \|\theta\|_\infty^8\right)$$

we get that $E \leq \delta_1$. Replacing E by its definition, this proves that:

$$E = q^\top \bar{w} - q^\top (w^\star/\varepsilon) + \lambda \max_{x\in\mathcal{X}} (h_x(\bar{w}))^+ \leq \delta_1$$

This allows to upper bound the constraints violation:

$$\lambda \max_{x\in\mathcal{X}} (h_x(\bar{w}))^+ \leq \frac{\delta_1 - q^\top \bar{w} + q^\top (w^\star/\varepsilon)}{\lambda} \leq \frac{\delta_1 + m^2 d \|\theta\|_\infty}{\lambda}$$

using the fact that $q^\top \bar{w} \geq 0$ and $q^\top w^\star \leq md\|\theta\|_\infty$ from Lemma 1. Setting

$$\lambda = \frac{1}{\delta_2}(\delta_1 + m^2 d \|\theta\|_\infty)$$

this proves that

$$\max_{x\in\mathcal{X}}(h_x(\bar{w}))^+ \leq \delta_2$$

Define $\bar{w}' = (1+\delta_2)\bar{w}$. Since, for all x,

$$\sum_{i\in\mathcal{I}} \frac{x_i}{\bar{w}_i} \leq \Delta_x^2 + \delta_2$$

we have that

$$\sum_{i\in\mathcal{I}} \frac{x_i}{\bar{w}'_i} \leq \frac{\Delta_x^2+\delta_2}{1+\delta_2} = \Delta_x^2 \frac{\frac{1}{\Delta_x^2}+\delta_2}{1+\delta_2} \leq \Delta_x^2$$

using the fact that $\Delta_x^2 \geq 1$ from Assumption 2. Hence, $\max_{x\in\mathcal{X}}(h_x(\bar{w}'))^+ = 0$, which means that $\bar{w}'$ is a feasible solution.

Finally:

$$q^\top \bar{w} - q^\top(w^\star/\varepsilon) \leq q^\top \bar{w} - q^\top(w^\star/\varepsilon) + \lambda \max_{x\in\mathcal{X}}(h_x(\bar{w}))^+ \leq \delta_1$$

so that

$$q^\top \bar{w}' - q^\top(w^\star/\varepsilon) \leq (1+\delta_2)\delta_1 + \delta_2 q^\top(w^\star/\varepsilon) \leq (1+\delta_2)\delta_1 + \delta_2 m^2 d\|\theta\|_\infty/\varepsilon = \frac{\delta}{2}+\frac{\delta}{2} = \delta.$$

Putting it all together, we have proven that $\bar{w}'$ is a feasible solution which verifies

$$q^\top \bar{w}' - q^\top(w^\star/\varepsilon) \leq \delta$$

This concludes the proof.

7.7 Proof of Proposition 4

We prove the result using recursion. Assume that $\bar{w}^k$ can be written as a linear combination with positive coefficients of elements of $\mathcal{X}$. Define $\mathcal{J}^k = \{i = 1, ..., d : w_i > 0\}$ the set of its non-null entries. We have $x^k \in \arg\min_{x\in\mathcal{X}} \sum_{i\notin\mathcal{J}} x_i$.

Since, by assumption, $\bar{w}^k$ can be written as a linear combination with positive coefficients of elements of $\mathcal{X}$, there exists x such that $\sum_{i\notin\mathcal{J}^k} x_i = 0$. Therefore, by definition of x^k, we must have $0 \leq \sum_{i\notin\mathcal{J}^k} x_i^k \leq \sum_{i\notin\mathcal{J}^k} x_i = 0$, so that $\sum_{i\notin\mathcal{J}} x_i^k = 0$. We then write

$$\bar{w}^{k+1} = \bar{w}^k - x^k \alpha_{x^k}$$

Now, we have that $\bar{w}^{k+1} \geq 0$ since $\bar{w}_i^{k+1} = \bar{w}_i^k$ if $i \notin \mathcal{J}^k$ and $\bar{w}_i^{k+1} = \bar{w}_i^k - \min_{i\in\mathcal{J}^k} \bar{w}_i^k \geq 0$ if $i \in \mathcal{J}^k$. Furthermore, we have that

$$M\bar{w}^{k+1} = \alpha_{x^k} M x^k + M\bar{w}^k = M\bar{w}^k = 0$$

since $x^k \in \mathbf{conv}(\mathcal{X})$, which implies $Mx^k = 0$. Therefore, $\bar{w}^{k+1}$ can be written as a linear combination with positive coefficients of elements of $\mathcal{X}$. Also, $|\mathcal{J}^{k+1}| \leq \max(0, |\mathcal{J}^k| - 1)$: indeed, we have that $\mathcal{J}^{k+1} \subset \mathcal{J}^k$ and, if $i_k \in \arg\min_{i\in\mathcal{J}^k} w_i^k$, we have that $w_{i_k}^k > 0$ and $w_{i_k}^{k+1} = 0$ by construction.

Since $w^1 = w^\star$ can be written as a linear combination with positive coefficients of elements of $\mathcal{X}$, the above argument shows that w^k can be written as a linear combination with positive coefficients of elements of $\mathcal{X}$ for all k, and that $w^{d+1} = 0$. This implies that the procedure does terminate after at most d iterations and

$$w^\star = \sum_{k=1}^{d} x^k \alpha_{x^k}$$

with $x^1, ..., x^d$ in $\mathcal{X}$ and $\alpha_{x^1}, ..., \alpha_{x^d}$ positive numbers.

For each iteration, it is noted that x^k can be computed by linear maximization over $\mathcal{X}$, which is feasible in time $O(\textbf{poly}(d))$ from Assumption 3. Since the above procedure terminates after at most d iterations, it takes $O(\textbf{poly}(d))$ time.

7.8 Projection Step

The projection of $w^t - \eta g^t$ on $\mathcal{M}$ involves solving the following optimization problem:

$$\begin{aligned} &\underset{w \in \mathbb{R}^d}{\text{minimize}} \ \|w - w^t - \eta g^t\|^2 \\ &\text{subject to } Mw = 0, w \geq 0, \min_{i \in \mathcal{I}} w_i \geq \underline{w}. \end{aligned}$$

There are two possible cases: *(i)* in some cases, this projection may be computed exactly, *(ii)* the projection may be computed using an interior point method using a logarithmic barrier function and a Newton step [3] Chapter 11.2. By definition, $w_t \in \mathcal{M}$, so that w_t can be used as an initial feasible point to compute the projection. We also mention that there exists even more efficient algorithms for specific combinatorial sets, for instance for the matching polytope [18].

Inputs: A and b (representation for the convex hull of $\mathcal{X}$), θ (mean reward vector), δ (accuracy level), ε (approximation ratio)
Parameter choice: set

$$\delta_2 = \frac{\delta\varepsilon}{m^2 d\|\theta\|_\infty}$$
$$\delta_1 = \frac{\delta}{2(1+\delta_2)}$$
$$\lambda = \frac{1}{\delta_2}(\delta_1 + m^2 d\|\theta\|_\infty)$$
$$T = \frac{1}{\delta_1^2}\varepsilon^{-2}m^5 d^2\|\theta\|_\infty^2\left(\|q\|^2 + \lambda^2\varepsilon^{-2}dm^8\|\theta\|_\infty^8\right)$$
$$\eta^2 = \frac{\varepsilon^{-2}m^5 d^2\|\theta\|_\infty^2}{T(\|q\|^2 + \lambda^2\varepsilon^{-2}dm^8\|\theta\|_\infty^8)}$$

Step 1: Dimensionality Reduction

Compute $M = A - \frac{bb^\top A}{\|b\|^2}$ and $q = (\theta^\top x^\star)\frac{b^\top A}{\|b\|^2} + \theta$ and $\underline{w} = (m\|\theta\|_\infty)^{-2}$

Step 2: Approximate Gradient Descent

Set $w^0 = (\underline{w}, ..., \underline{w})$
For $t = 1, ..., T$:
Find x^t such that $\max_{x\in\mathcal{X}} h_x(w^t) \le h_{x_t}(\varepsilon w^t)$
Compute $g^t = q + \lambda\varepsilon\nabla h_{x^t}(\varepsilon w^t)\mathbf{1}(h_{x^t}(\varepsilon w^t) > 0)$
Update $w^{t+1} = \Pi_{\mathcal{M}}\left\{w^t - \eta g^t\right\}$
Compute $\bar{w} = \frac{1}{T}\sum_{t=1}^T w_t$ and $\bar{w}' = (1+\delta_2)\bar{w}$

Step 3: Retrieving the Solution to the Original Problem

Set $\bar{w}^1 = \bar{w}'$
For $k = 1, ..., d$:
Find $x^k \in \arg\min_{x\in\mathcal{X}} \sum_{i=1}^d x_i \mathbf{1}\{\bar{w}_i^k > 0\}$
If $\bar{w}^k > 0$ let $\alpha_{x^k} = \min_{i:\bar{w}_i^k>0} \bar{w}_i^k$, otherwise let $\alpha_{x^k} = 0$.
Update $\bar{w}^{k+1} = \bar{w}^k - \alpha_{x^k}x^k$.
Output: A (δ, ε)-optimal solution to the Graves Lai optimization problem $\alpha_{x^1}, ..., \alpha_{x^d}$ and $x^1, ..., x^d$.

Figure 2: The GLPG Algorithm: Computing the Solution to P_{GL} in Polynomial Time.